Nuclear phylogenomics, but not mitogenomics, resolves the most successful Late Miocene radiation of African mammals (Rodentia: Muridae: Arvicanthini)

Nuclear phylogenomics, but not mitogenomics, resolves the most successful Late Miocene radiation of African mammals (Rodentia: Muridae: Arvicanthini)

Ondřej Mikula [a,b], Violaine Nicolas [c], Radim Šumbera [d], Adam Konečný [e], Christiane Denys [c], Erik Verheyen [f,g], Anna Bryjová [a], Alan R. Lemmon [h], Emily Moriarty Lemmon [i], and Josef Bryja [a,e,*]

[a] Institute of Vertebrate Biology of the Czech Academy of Sciences, 603 65, Brno, Czech Republic

[b] Institute of Animal Physiology and Genetics of the Czech Academy of Sciences, 602 00, Brno, Czech Republic

[c] Institut de Systématique, Evolution, Biodiversité (ISYEB), Muséum national d'Histoire naturelle, CNRS, Sorbonne Université, EPHE , Université des Antilles, CP51, 75005 Paris, France;

[d] Department of Zoology, Faculty of Science, University of South Bohemia, 370 05, České Budějovice, Czech Republic

[e] Department of Botany and Zoology, Faculty of Science, Masaryk University, 611 37, Brno, Czech Republic

[f] Royal Belgian Institute for Natural Sciences, Operational Direction Taxonomy and Phylogeny, 1000, Brussels, Belgium

[g] Evolutionary Ecology Group, Biology Department, University of Antwerp, 2020, Antwerp, Belgium

[h] Department of Scientific Computing, Florida State University, Dirac Science Library, Tallahassee, FL 32306-4295, United States

[i] Department of Biological Science, Florida State University, 319 Stadium Drive, PO Box 3064295, Tallahassee, FL 32306-4295, United States

***Correspondence**: Josef Bryja, Institute of Vertebrate Biology of the Czech Academy of Sciences, Research Facility Studenec, Studenec 122, 675 02 Koněšín, Czech Republic; E-mail: bryja@ivb.cz

***Running head*:** Phylogenomics of Arvicanthini rodents

ABSTRACT

The tribe Arvicanthini (Muridae: Murinae) is a highly diversified group of rodents (ca. 100 species) and with 18 African genera (plus one Asiatic) represents probably the most successful adaptive radiation of extant mammals in Africa. They colonized a broad spectrum of habitats (from rainforests to semi-deserts) in whole sub-Saharan Africa and their members often belong to most abundant parts of mammal communities. Despite intensive efforts, the phylogenetic relationships among major lineages (i.e. genera) remained obscured, which was likely caused by the intensive radiation of the group, dated to the Late Miocene. Here we used genomic scale data (377 nuclear loci; 581,030 bp) and produced the first fully resolved species tree containing all currently delimited genera of the tribe. Mitogenomes were also extracted, and while the results were largely congruent, there was less resolution at basal nodes of the mitochondrial phylogeny. Using a newly developed algorithm for subsampling of most informative loci, we also performed a fossil-based divergence dating. The results suggest that the African radiation started early after the colonization of Africa by a single arvicanthine ancestor from Asia during the Messinian stage (ca. 7 Ma), and was likely linked with a fragmentation of the pan-African Miocene forest. Some lineages remained in the rain forest, while many others successfully colonized broad spectrum of new open habitats (e.g. savannas, wetlands or montane moorlands) that appeared at the beginning of Pliocene. One lineage even evolved partially arboricolous life style in savanna woodlands, which allowed them to re-colonize equatorial forests. We also discuss delimitation of genera in Arvicanthini and propose corresponding taxonomic changes.

Keywords: Late Miocene, radiation, anchored phylogenomics, Rodentia, tropical Africa, complete mitochondrial DNA

1. Introduction

The murid rodents (Rodentia: Muridae) are evolutionarily the most successful group of mammals in the Old World, with 816 currently recognized species (Wilson et al., 2017). Their phylogeny is relatively well known thanks to recent analyses of large multi-locus genetic datasets and calibration of a molecular clock based on multiple paleontological records (e.g. Steppan and Schenk, 2017; Aghová et al., 2018). Among five subfamilies, Murinae form the majority of murid rodents (ca. 80%; Wilson et al., 2017). They evolved in *ca* 15 major clades (= tribes) (Steppan and Schenk, 2017) with very unequal distribution of species diversity (a single species in Micromyini vs. 185 species in Rattini; Wilson et al., 2017). Five murine tribes (Otomyini, Arvicanthini, Malacomyini, Murini, Praomyini) are indigenous in sub-Saharan Africa (Lecompte et al., 2008) and they constitute the most species-rich group of African mammals.

The tribe Arvicanthini (Lecompte et al., 2008; Denys et al., 2017) is the most speciose tribe of African rodents with 18 currently recognized extant African genera and the Asiatic genus *Golunda* (Denys et al., 2017; Missoup et al., 2018; Table S1 in SM1). Some genera are species-rich and widely distributed (e.g. *Lemniscomys* - 11 species, *Aethomys* - 9 species, *Grammomys* - 14 species), while others have low diversity and restricted ranges, e.g. one species of *Lamottemys* from Mt. Oku in Cameroon, or two species of *Desmomys* from Ethiopian highlands (Denys et al. 2017). They colonized the whole of sub-Saharan Africa (two species have also isolated populations in Maghreb and Egypt), where they live in a broad spectrum of habitats; from lowland and montane rainforests through various types of open habitats (marshlands, savannas, woodlands) to semi-deserts. The first radiation of the tribe occurred in Late Miocene after arrival of arvicanthine ancestor(s) from Asia (7-9 Mya; Aghová et al. 2018) and was likely related to intensive climatic changes and a spread of open habitats (Lecompte et al., 2008). Most modern genera of

Arvicanthini appeared almost simultaneously and they can serve as a model for understanding the evolutionary process of (adaptive) radiation.

A reliable phylogenetic reconstruction is required for deciphering mechanisms of such successful radiation. However, despite intensive efforts, the phylogenetic relationships among many genera of Arvicanthini are still uncertain. Previous studies employed mitochondrial (Ducroz et al., 2001) or the combination of a limited number of mitochondrial and nuclear sequences (Lecompte et al., 2008; Missoup et al., 2016; Bryja et al., 2017; Steppan and Schenk, 2017; Aghová et al., 2018; Missoup et al., 2018). Even if these studies agreed e.g. on the monophyly of the so-called *Hybomys* division (sensu Musser and Carleton, 2005) or the sister relationship of *Lemniscomys* and *Arvicanthis*, numerous (especially deeper) nodes on the phylogenetic tree remained unresolved. They are either unsupported or have conflicting topologies dependent on the markers used, which may be the outcome of the rapid radiation of the Arvicanthini in Late Miocene (Aghová et al., 2018).

Increasing the amount of genetic data frequently allows resolution of even the most problematic phylogenetic relationships. One such approach is based on sequencing of complete mitochondrial genomes ("mitogenomics"), instead of single mitochondrial genes; this helped to reconstruct e.g. the phylogeny of primates (Pozzi et al., 2014) or sharks and rays (Amaral et al., 2018). However, because of the absence of recombination, the mitochondrial DNA should be still considered as a single locus and reconstructed phylogenies represent only "single gene" trees. To address this problem, recent phylogenomic approaches target markers derived from moderately conserved regions, mostly exons and surrounding introns (Lemmon et al., 2012), or ultraconserved genomic elements and their flanking regions (McCormack et al., 2012), which allow to infer a species-tree

that accounts for discord among hundreds of independent loci at nuclear DNA (Lemmon and Lemmon, 2013). These regions are enriched in genomic libraries by hybridization and then sequenced by high-throughput sequencing. They can be analysed even from old museum material (e.g. McCormack et al., 2016) and they allowed solving the notoriously difficult nodes in phylogeny of birds (Prum et al., 2015) or placental mammals (McCormack et al., 2012).

Here we used the so-called anchored phylogenomic approach (Lemmon et al., 2012) to infer the most reliable phylogenetic tree for the Arvicanthini. This is the first multi-locus analysis including all extant African genera of this clade, as well as the Asian genus *Golunda*. With the resolved topology of the tribe in hand, we estimated the time-frame, during which this tribe radiated and assessed its evolutionary history in the context of environmental changes since Late Miocene. As a by-product of sequencing of anchored loci, we assembled also complete mtDNA from all samples and we compared the ability of anchored phylogenomics vs. mitogenomics in phylogenetic reconstruction of a fast mammalian radiation.

2. Material and methods

2.1 Taxon sampling

The final dataset analysed in this study includes 40 genotyped specimens (= one individual per species; Table S1) representing all 18 nominal genera of the African Arvicanthini, as well as the closely related Asian genus *Golunda* belonging to the same tribe (Denys et al., 2017; Missoup et al., 2018). Two species of the tribe Otomyini and one of Millardiini were chosen as the closest relatives of Arvicanthini and two species of the tribe Praomyini were used as more distant outgroups within the subfamily Murinae (Lecompte et al., 2008; Aghová et al., 2018) (Table S1).

2.2 Anchored hybrid enrichment (AHE) data collection and assembly of nuclear dataset

Probe design and data collection were performed by the Center for Anchored Phylogenomics (www.anchoredphylogeny.com). Following Ruane et al. (2015; snakes), Tucker et al. (2016; lizards), and Prum et al. (2015; birds), we improved the vertebrate AHE target loci of Lemmon et al. (2012) for optimal use in mammals. We first identified the genomic coordinates in the human genome (hg19) corresponding to the coordinates of the extended anchor regions of *Gallus gallus* (galGal4) obtained by Prum et al. (2015) using the UCSC liftover tool (http://genome.ucsc.edu/cgi-bin/hgLiftOver). The corresponding genomic sequences were then extracted and aligned using MAFFT v7.023b (Katoh and Standley 2013) to that of the regions used by Prum et al. (2015) for probe design. After inspecting the alignments and masking any misaligned regions in Geneious R9 (Biomatters Ltd.; Kearse et al. 2012), 120 bp probes were tiled uniformly across the human sequences at 1.5x density. Genomic DNA was extracted using the Invisorb® Spin Tissue Mini Kit (Stratec, Germany). After extraction, indexed libraries were prepared on a Beckman Coulter FXP liquid-handling robot following Lemmon et al. (2012) and Prum et al. (2015). Libraries were then pooled at equal concentrations in three groups of ~14 samples and enriched using an Agilent SureSelect XT kit containing the probes described above. Enriched library pools were then sequenced on one paired-end 150 bp lane (43 Gb of raw data) of an Illumina HiSeq 2500 sequencer at the Translational lab in the Florida State University.

In order to increase read accuracy and length, paired reads were merged prior to assembly following Rokyta et al. (2012), which also removes adapter sequences. Following the approaches of Prum et al. (2015) and Hamilton et al. (2016), a quasi-de novo assembly approach was taken using *Homo sapiens* as the reference. Assembly clusters derived from fewer than 175 reads were removed from further analysis in order to reduce the possible effects of low level contamination

and mis-indexing. Orthology was established among the consensus sequences recovered at each of the target loci using the pairwise sequence distances among the consensus sequences (see Hamilton et al., 2016 for details). After orthologous sequences were then aligned using MAFFT v7.023b (Katoh and Standley 2013; with --genafpair and --maxiterate 1000 flags utilized), the alignments were trimmed/masked to remove poorly aligned regions (following Hamilton et al. 2016; with the following parameters: MinGoodSites=14, MinPropSame=0.4, and MissingAllowed=20). Finally, trimmed alignments were inspected in Geneious and any remaining misaligned regions were masked.

2.3 Assembly and alignment of mitogenomes

Mitochondrial DNA is usually highly prevalent in genomic DNA extractions and it still persists even in genomic libraries enriched for particular conserved loci. As a by-product of AHE approach, we therefore used the raw data of Illumina reads to assembly the complete mitogenomes of 40 analysed taxa. Heavy-strand protein-coding genes (12 genes) and genes for non-coding RNA (two ribosomal RNAs and 22 transfer RNAs) were extracted from the complete mitochondrial sequences in Geneious according to the annotated references of complete mtDNAs of *Apodemus draco* (GenBank accession number KP694301) and *A. chevrieri* (HQ896683) from the relatively closely related tribe Apodemini (Murinae). Following Pozzi et al. (2014), we excluded the D-loop sequences because of alignment difficulties (highly variable non-coding sequences), and ND6 gene because it is encoded on the mitochondrial L-strand which has a different nucleotide composition from the H-strand, and has been shown to have poor phylogenetic signal (Gissi et al., 2000). Protein-coding genes were individually aligned based on their corresponding amino acid translations using Muscle 3.8 (Edgar, 2004) implemented in AliView 1.18 (Larsson, 2014). Two genes for ribosomal RNA (12S-rDNA and 16S-rDNA) and 22 genes for transfer RNA were aligned

separately by online version of MAFFT v. 7 (http://mafft.cbrc.jp/alignment/server/) using the algorithm Q-INS-i, which considers secondary structure of RNA and is recommended for a global alignment of highly divergent non-coding RNAs (Katoh and Toh, 2008). The resulting alignments of genes for both rRNAs and tRNAs were analysed by Gblocks 0.91b (Castreana, 2000). Gblocks removes all poorly aligned regions in a dataset, which has been shown to be particularly effective in phylogenetic studies including very divergent sequences (Talavera and Castreana, 2007). Gblocks was run with the options "Minimum Length Of A Block" = 5, and "Allowed Gap Positions" = "With Half".

2.4 Phylogenetic analysis of nuclear loci - species tree in ASTRAL

Multispecies coalescent (MSC) provides sound foundation for species tree inference as it models incomplete lineage sorting and hence discordance between gene trees (Degnan and Rosenberg, 2009). However, joint estimation of species tree and gene trees becomes too computation expensive with large numbers of loci (Ogilvie et al., 2017). For this reason we inferred species tree by ASTRAL II (v. 4.11.2, Mirarab and Warnow, 2015) – a summary method analysing topologies of pre-estimated gene trees by breaking them into a multi-set of quartet trees and searching for species tree inducing quartet tree topologies that are most frequently observed in the multi-set (Mirarab and Warnow, 2015). The gene trees were obtained in separate Bayesian analyses using MrBayes v. 3.2.6 (Ronquist et al., 2012). They were inferred as unrooted with uniform prior probability over tree topologies. Branch lengths were unconstrained by clock assumptions and we used exponential prior (μ=10.0) for each of them. Integral to the analysis was sampling of time reversible nucleotide substitution models (Huelsenbeck et al., 2004) by reversible jump Markov Chain Monte Carlo. Gamma-distributed rate variation (discretized into eight categories) was assumed among sites. The template of MrBayes block in the *nexus* file is available as SM2. ASTRAL

accepts just a single tree per gene and thus it was necessary to find a tree representing the whole posterior sample obtained from MrBayes. It was defined as a maximum bipartition credibility tree (MBCT), i.e. the tree with maximum product of its bipartitions' posterior probabilities (cf. Drummond and Bouckaert, 2015, p. 94). Branch lengths are not used in ASTRAL and thus only MBCT topology was calculated in package 'phangorn' (Schliep, 2011) for R (R Core Team, 2019). In general, the gene trees were not fully resolved and the poorly supported bipartitions could mislead ASTRAL. Therefore, the bipartitions with posterior probability (PP) < 0.90 were collapsed, creating a polytomy in the tree. Calculation of PPs was done in 'ape' (Paradis and Schliep, 2018).

2.5 Bayesian phylogeny of mitogenomes

We used PartitionFinder v. 2 (Lanfear et al., 2016) to simultaneously detect partitions and the most suitable substitution models for different parts of mtDNA. Using AICc criterion, the best scheme supported 12 partitions (the partitioned *nexus* input file is in SM3). Bayesian analysis of evolutionary relationships was performed in MrBayes v. 3.2.6, employing Markov Chain Monte Carlo (MCMC) simulations of posterior probability. Three heated and one cold chain were employed in an analysis with 12 partitions, and runs were initiated from random trees. Two independent runs were conducted with 20 million generations each and trees and parameters were sampled every 1000 generations. Convergence was checked using TRACER v1.5 (Rambaut and Drummond, 2007). For each run, the first 20% of sampled trees were discarded as burn-in. Bayesian posterior probabilities were used to assess branch support of the maximum clade credibility tree with common ancestor node heights.

2.6 Maximum likelihood estimation of mtDNA and nuclear phylogenies

The statistical methods used here for the species tree and mitochondrial tree inference are computationally demanding and we therefore applied also complementary and much faster maximum likelihood inference in RAxML 8.2.10 (Stamatakis, 2014). The nuclear and mitochondrial datasets were analysed separately. Individual loci in both datasets were concatenated and hence assumed to share the same phylogeny, which is realistic only in physically linked mitochondrial loci, but not in unlinked nuclear loci. Because simpler models are not available in RAxML, the GTR+G model was used for all partitions, which were allowed to differ in their substitution parameters. For mtDNA the partitions were defined as described above and in nuclear data set every locus corresponded to a single partition. The robustness of the nodes was evaluated by the rapid bootstrap procedure (Stamatakis et al., 2008) with 1000 replications.

2.7 Time-calibrated phylogeny

The time-calibrated history of divergences between arvicanthine species was inferred in StarBEAST2 (Ogilvie et al., 2017). The species tree topology as well as gene tree topologies were fixed to the estimates obtained by ASTRAL and MrBayes, respectively, but the branch lengths were allowed to vary. We assumed species tree to arise in a constant rate birth-death process (Gernhard, 2008) with uninformative priors put on its parameters. Outgroups used to root the trees were excluded in this analysis. Time information was injected into the species tree by two fossil-based constraints on ages of specific ancestral nodes. Firstly, 9.2 million years (Ma) old †*Karnimata darwini* (Kimura et al., 2015) constrained the age of the root, i.e. the most recent common ancestor (MRCA) of Millardiini/Otomyini/Arvicanthini clade. Secondly, †*Aethomys* sp. and †*Arvicanthis* sp. fossils from 6.1 Ma old site Lemudong-o' (Manthi, 2007) constrained the age of MRCA of *Aethomys* and *Arvicanthis*. The fossils were taken from the set proposed for subfamily Murinae by Aghová et al. (2018), but the latter two were used more conservatively to account for

a possibility they represent just members of lineages leading to the particular genera. The calibration densities were uniform: 9.2-11.2 Ma for the root and 6.1-11.2 for *Aethomys*/*Arvicanthis*. The maximum age 11.2 Ma was motivated by the fossil of †cf. *Karnimata* from Nagri Formation, Siwalik Group, interpreted to be close to the split of lineages leading to extant *Mus* and *Arvicanthis* (Kimura et al., 2015; Aghová et al., 2018).

For the time calibration analysis we considered just 269 out of 377 nuclear loci, namely those successfully sequenced in all species and not having outgroup and ingroup species intermixed in single-gene topologies. In each of these loci we compared strict and uncorrelated lognormal relaxed clock (Drummond et al., 2006) using Bayes factors calculated in RevBayes v. 1.1.0 (Höhna et al., 2016). For the analysis, we retained 231 loci where the strict clock model was supported. Nucleotide substitution model parameters were fixed to the averages of posterior samples obtained from the MrBayes analyses.

Two independent runs of the analysis were conducted to check for convergence. The pooled posterior sample was represented by the Maximum Clade Credibility (MCC) tree with the mean common ancestor node heights (Drummond and Bouckaert, 2015).

3. Results

3.1 Summary of collected data sets

The nuclear phylogenomic analysis was based on 377 successfully sequenced loci ranging in length from 436 to 2,565 bp (median 1,644 bp). The total length of concatenated alignment for 40 taxa was 581,030 bp. Some sequences were incomplete or missing and thus the data set contained from 22 to 40 sequences for particular loci. Overall, 3.9% sequences and 5.6% bp were missing.

For the same 40 individuals we produced the unambiguous mtDNA alignments for 12 protein-coding genes (10,891 bp), two rRNA-coding genes (2,419 bp) and 22 tRNA-coding genes (1,467 bp). These alignments were concatenated into final mitogenomic alignment of 14,777 bp, equivalent to approximately 91% of the rodent mitochondrial genome.

3.2 Phylogenetic reconstructions based on multilocus nuclear data and complete mtDNA

The results of phylogenetic analyses are summarized in Fig. 1. After rooting by Praomyini, the nuclear species tree shows *Millardia meltada* as the sister lineage to Otomyini+Arvicanthini and hence the tribe Otomyini as the sister lineage of Arvicanthini. The only living Asian species of Arvicanthini, *Golunda ellioti*, is in sister relationship to all African taxa. In Africa, the basal split is between *Oenomys* and remaining genera, where we recognize four major clades, named here *Hybomys*, *Aethomys*, *Dasymys* and *Arvicanthis* clades (Fig. 1, Table S1). The phylogeny is almost fully resolved (PP=1.00), just the position of *Thallomys* has slightly lower support (PP=0.92). The same topology was obtained from the ML analysis of concatenated loci and also the bootstrap support (BS) was maximum for all nodes, except for *Thallomys–Thamnomys*/*Grammomys* node with BS=96%.

On the contrary, the phylogenetic tree based on mitochondrial genomes shows low support for relationships between *Hybomys*, *Aethomys*, *Dasymys* and *Arvicanthis* clades, as well as within the *Aethomys* clade. The topology is generally similar to the nuclear tree, but differs in the following points: (i) *Golunda* and *Oenomys* form a strongly supported (PP=1.00) monophyletic clade rather than subsequent offshoots, (ii) *Rhabdomys* is supported (PP=0.94, BS=98%) as the sister of *Lamottemys* rather than of *Desmomys*, (iii) there are two differences in topology within genera *Arvicanthis* and *Lemniscomys*. For each of these conflicts we examined the number of MrBayes

gene trees, whose topology was congruent with either nuclear or mitochondrial tree. The position of *Oenomys* on gene trees varied considerably. In equal share of 9% gene trees *Oenomys* was sister to "non-*Golunda* Arvicanthini" (nuclear topology), *Golunda* (mitochondrial topology) or to the rest of arvicanthini, but in smaller proportions of gene trees it was found sister to many different clades. *Rhabdomys* was sister to *Desmomys* (nuclear topology) in 40% of gene trees, while in 21% it was sister to *Lamottemys* (mitochondrial topology). Within *Arvicanthis* and *Lemniscomys*, the percentage was 41% to 26% and 61% to 18%, respectively, always in favour of the nuclear topology.

3.3 Diversification in the historical context

The split between the tribe Arvicanthini and its sister tribe Otomyini is dated to the Tortonian stage of Late Miocene (9.0 Ma) (Fig. 2). Still in the same stage, the Asian *Golunda* diverged from the ancestor of all African arvicanthines (7.9 Ma). The most intensive African radiation, when a majority of modern genera appeared, overlaps with the Messinian stage of Late Miocene (the times of most recent ancestors, TMRCAs, clustered in the period between 7.6 and 5.3 Ma). The next intensive diversification period is dated to Lower-Pliocene (4.7 - 3.7 Ma) with intergeneric splits between *Stochomys*/*Dephomys*, *Grammomys poensis*/*Thamnomys*, *Lamottemys*/*Desmomys*/*Rhabdomys* and the oldest diversifications within the *Arvicanthis* clade. The first intrageneric divergences (in *Aethomys*) are dated in the same period. The oldest splits within other genera (*Typomys*, *Grammomys*) are dated to the end of Pliocene, which overlaps with the divergence between the youngest genera of Arvicanthini, i.e. *Pelomys*/*Mylomys* and *Arvicanthis*/*Lemniscomys* (Fig. 2).

4. Discussion

4.1 Anchored phylogenomics vs. complete mtDNA

We demonstrated that complete mtDNA is less powerful and less reliable in resolving phylogenetic relationships than the anchored nuclear loci; this resulted in lack of resolution of some of the deep nodes (dated to 6.3-7.6 Ma) in the Bayesian analysis of complete mtDNA. This is probably due to higher substitution rates in mtDNA, which makes it largely saturated by mutations on larger timescales. Also, mtDNA was much shorter (14,777 bp compared to 581,030 bp) and its analysis cannot benefit from modelling of gene tree discordance and its potential to bring additional information about phylogenetic relationships at the species level. Finally, variation in mtDNA may be more affected by selection due to prevalence of coding sequences.

There are also some differences in topology of mtDNA and nuclear trees. The relationships inferred by anchored phylogenomics have higher credit here because mtDNA tree may differ from the species tree due to incomplete lineage sorting and, especially on shallow scales (e.g. within *Arvicanthis*, *Lemniscomys*), also due to mitochondrial introgression. Notably, the incongruent nodes are usually poorly supported by at least one phylogenetic method at mitochondrial tree. Gene trees were predominantly congruent with the nuclear species tree topology dominated in all but one conflicting relationships. The exception was *Oenomys*, whose position in gene trees was very variable and its confidential placement by ASTRAL was apparently driven by distribution of quartet subtrees rather than by prevalence of fully congruent bipartitions in input unrooted trees. The subsequent dating analysis estimated close to zero branch length separating *Oenomys* from non-*Golunda* arvicanthines and so hard polytomy may be suspected here. Taken together, we will base the following discussion on the species tree obtained by anchored phylogenomics.

4.2 Evolutionary origin of Arvicanthini

The tribe Arvicanthini (Denys et al., 2017) forms a strongly supported monophyletic group, sister to Otomyini and more distantly to Millardiini. The monophyly of the tribe has been repeatedly recognized in previous phylogenetic analyses based on few genetic segments of mitochondrial and nuclear DNA (Ducroz et al., 2001; Steppan et al., 2004, 2005; Lavrenchenko and Verheyen, 2005; Lecompte et al., 2008; Rowe et al., 2008; Schenk et al., 2013; Missoup et al., 2016, 2018; Steppan and Schenk, 2017; Aghová et al., 2018), although no multi-locus genetic study has integrated all nominal genera of the tribe; even the most complete study of Missoup et al. (2018) used only mtDNA for *Thamnomys*. Our phylogenomic analysis with complete sampling of all genera not only confirms monophyly of Arvicanthini, but for the first time fully resolves phylogenetic relationships among genera within the tribe, which now allows the reconstruction of their (adaptive) radiation.

All recent African genera of Arvicanthini form a monophyletic group, sister of Indian *Golunda*. The basal position of *Golunda* in Arvicanthini was for the first time suggested by Lecompte et al. (2008), while in many other multi-locus phylogenies, *Golunda* formed a sister group of *Oenomys* (e.g. Steppan and Schenk, 2017; Aghová et al., 2018). The latter topology was likely affected by mtDNA variation, because sister relationship of the two genera was revealed also in our mitogenomic phylogeny (Fig. 1). The successive sisters of Arvicanthini (with one Indian and one African lineage) are African Otomyini and Indian Millardiini. It seems therefore reasonable to ask whether these rodents diversified in Asia and then twice colonized Africa by ancestors of Otomyini and African Arvicanthini or vice versa (i.e. diversification of the group in Africa and re-colonization of Asia by *Golunda* and the ancestor of Millardiini). The origin of murine rodents, subfamily Murinae, in Asia is well supported (e.g. Aghová et al., 2008; Schenk et al., 2013; and references therein). Lecompte et al. (2008) pointed out that multiple sister relationships between Asian and

African clades (Praomyini - Murini; Malacomyini - Apodemini; Arvicanthini/Otomyini - Millardiini, respectively) suggest that each of the African lineages was differentiated prior to their dispersal into Africa. This should happened around the same time as a part of broader episode of faunal interchange (11-10 Ma), which is in a very good fit with the fossil record (see references in Lecompte et al., 2008, and review in Winkler et al., 2010). Applying this logic to the resolved phylogeny of Arvicanthini, we can speculate that even the ancestors of African Arvicanthini and Otomyini diverged already in Asia and arrived to Africa independently in Late Miocene (ca. 9-7 Ma), leaving *Golunda* (as the only surviving lineage of already differentiated Arvicanthini) in Asia.

Palaeontological records, however, do not refute an alternative explanation, making the discussion about African or Asian origin of Arvicanthini even more complex. During the Middle Miocene (16.0-11.6 Ma), the Mediterranean-Indo-Pacific seaway closed again at the beginning of the Serravallian ca. 13.8 Ma ('Parathethys Salinity Crisis'; Rögl, 1999), and the newly formed land bridge allowed repeated exchanges of terrestrial organisms. We can hypothesize that ancestors of contemporary Arvicanthini occurred more or less continuously across this part of Afro-Asia. Murine fossil records provide clear evidence for connections between the Indomalaya, the Palearctic and the Afrotropics in this period (see references in Aghová et al., 2018). Among them, the conspicuous examples are the oldest records of †*Parapelomys* spp., considered to belong to Arvicanthini (Denys et al., 2017), found synchronously in Africa (8.5 Ma in Chorora, Ethiopia; Geraads, 2001) and in Pakistan (ca. 8.0 Ma; Jacobs and Flynn, 2005). In agreement with that, the Ethiopian as well as Moroccan sites of early Pliocene epoch reportedly contain fossils identified as *Millardia* and *Golunda* (e.g. WoldeGabriel et al., 1994; Wynn et al., 2006), currently limited to the Indian subcontinent. The hypothesized sister genus of *Golunda*, †*Saidomys*, was in late Miocene also widely distributed in both northern Africa and Asia (Winkler, 2002; Patnaik, 2014).

All these records suggest that we cannot simply define the place of diversification of *Golunda* and other arvicanthines and the direction of colonization. *Golunda* (but similarly also *Millardia*) may be just viewed as phylogenetic relict of a Miocene faunal interchange that have disappeared from Africa, while all other ancient Arvicanthini (†*Parapelomys*, †*Saidomys*) went to the extinction globally (Lecompte et al., 2008). Fossils already assignable to extant African arvicanthine genera date from Late Miocene through Early Pliocene, around 7-5 Ma (e.g. *Aethomys* or *Arvicanthis*, see review in Winkler et al., 2010 and Denys and Winkler 2015). This is in good agreement with our molecular dating, which suggests that the first radiation of African Arvicanthini occurred in the Messinian epoch of Late Miocene, when major lineages within this clade have already diversified.

4.3 Ancestral traits, mechanisms of (adaptive) radiation

The tribe Arvicanthini, with its high number of species and diverse ecological adaptations and lifestyles represents the most successful radiation of rodents in the African continent. African Arvicanthini are monophyletic, suggesting that the radiation started from a single ancestor lineage. The first two offshoots are represented by the genus *Oenomys* and the strongly supported *Hybomys* clade. All species from these two clades inhabit Guineo-Congolian zone, suggesting that the radiation of the tribe in Africa started in a forest. African forests developed by the late Cretaceous, and during the Middle Miocene climatic optimum they extended coast to coast across the equatorial zone (Maley, 1996; Morley and Kingdon, 2013). It is therefore likely that murine newcomers from Asia in late Miocene first adapted (or already came adapted) to this most widespread ecosystem. Two other tribes of murine rodents that entered Africa also in Late Miocene are either restricted to Guineo-Congolian forests (Malacomyini) or have there the highest evolutionary diversity (Praomyini), which provides additional support for this hypothesis (Aghová et al., 2018; our unpubl. genomic data).

The most characteristic features of the climate in tropical Africa since the Late Miocene is its increasing variability and overall aridification (Ségalen et al., 2007; Potts, 2013). The shrinking of forests was linked to development of more open savanna-like ecosystems, evidenced e.g. by spread of C4 grasses. They appear in East Africa in the Mid-Late Miocene and between 8 and 6 Ma they already represent significant part of diet of grazing mammals (Cerling et al., 1997). Arvicanthine rodents were among the most successful mammals (together with large ungulates) that colonized these newly emerging ecological niches. It is especially true for the *Arvicanthis* clade, whose members often dominate small mammal assemblages in the non-forested sub-Saharan habitats. First, the genera *Arvicanthis* and *Lemniscomys* are closely related with numerous common morphological, ecological and behavioural traits (Ducroz et al., 2001). Their TMRCA was estimated to 2.9 Ma, which is much more recent than previously thought (Aghová et al., 2018). They are widespread and often very abundant in various savannas, except South Africa. The second subclade includes *Mylomys* and *Pelomys*, poorly known taxa preferring open moist habitats in the forest-savanna mosaic, mostly along the equator (Wilson et al 2017). Finally, the third subclade includes *Desmomys*, *Lamottemys* and *Rhabdomys* (in agreement with Ducroz et al., 2001; Lavrenchenko and Verheyen, 2005; Lecompte et al., 2008; Missoup et al., 2016). The composition of this cluster seems rather surprising at first because *Desmomys* is endemic to Ethiopian highlands, while *Lamottemys* is endemic to Mount Oku in Cameroon and *Rhabdomys* is widespread in various open habitats of Eastern and Southern Africa. However, they share multiple morphological traits and all of them prefer relatively humid montane habitats (e.g. Ducroz et al., 2001; Denys et al., 2014; Missoup et al., 2016). It is only south of the Limpopo river that *Rhabdomys* becomes widespread in arid habitats, a fact that might be related to the absence of competition from *Arvicanthis* and *Lemniscomys* (Ducroz et al., 2001). The distribution pattern in

this subclade reinforces the hypothesis of recurrent connections between western and eastern African mountains in mid-Pliocene i.e. the period that was characterized by warm and wet climate in Africa (Feakins and deMenocal, 2010). As a consequence, the moist montane environments expanded and facilitated a trans-continental dispersal of their inhabitants (see Taylor et al. 2014 for another example in rodents). We showed for the first time (contrary to Missoup et al. 2016) that *Lamottemys* diverged first in this subclade (ca. 4.0 Ma), which is understandable given the geographical distance between eastern and western African mountains, and in agreement with the pollen records indicating an abrupt change in forest cover ca. 3.3 Ma (Bonnefille et al., 2004). The genera *Rhabdomys* and *Desmomys* are ecological vicariants in East African mountains (Wilson et al., 2017) and they split soon after the divergence of *Lamottemys*. Recent phylogenetic studies repeatedly suggested that Ethiopian highlands served as a source from which other eastern African mountains were colonized through forest corridors that were predicted on both sides of the Great Rift Valley during humid periods of Plio-Pleistocene (e.g. Bryja et al., 2014; Šumbera et al., 2018; Krásová et al., 2019). It is therefore easy to imagine that the ancestor of *Desmomys* and *Rhabdomys* inhabited a large north-south belt of East African montane grasslands and moorlands and the two genera definitely diverged after the end of warm and wet mid-Pliocene period.

As already suggested by the tree topologies in recent multi-locus phylogenetic studies (e.g. Schenk et al., 2013; Steppan and Schenk, 2017; Missoup et al., 2018; Aghová et al., 2018), *Dasymys* is the sister to the *Arvicanthis* clade. The genus *Dasymys* has semi-aquatic habits and is specialized to live in marshlands (Wilson et al 2017), which again supports the hypothesis that the origin of the speciose *Arvicanthis* clade is in relatively moist open habitats, which are still occupied by some *Rhabdomys*, *Pelomys*, *Mylomys* or *Desmomys* species. On the other hand, the adaptation to arid (sometimes even semi-desert) environments in some species of *Arvicanthis* and *Lemniscomys* and

southern African *Rhabdomys* is probably a trait that evolved more recently (in Pleistocene) in response to the increasing aridity (deMenocal, 1995) and allowed these groups to occupy large areas in both South and North of equator forest blocks.

One of the most important results of this study is the first unequivocal resolution of phylogenetic relationships of notoriously difficult taxa *Micaelamys*, *Aethomys*, *Thamnomys* and *Thallomys*. Phylogenomic analysis clearly shows that all these genera form together a monophyletic clade with *Grammomys* (named the *Aethomys* clade here, following "*Aethomys* division" of Musser and Carleton (2005)). While two basal offshoots of this clade (i.e. *Micaelamys* and *Aethomys*) consist of taxa with predominantly terrestrial activity in savannas, the three remaining genera form a monophyletic group of at least partially arboricolous taxa (*Thallomys*, *Thamnomys*, *Grammomys*). This suggests that the (pre-)adaptation to climb trees evolved only once in the common ancestor of these genera, already in Late Miocene and probably in woodlands of south-eastern Africa, where most species of *Thallomys* and *Grammomys* are found today and where is also the highest diversity of *Aethomys*/*Micaelamys*. Based on the topology of the phylogenomic tree we can even speculate that the ability to climb - to gain access to resources that are above the ground and to protect themselves against predators - was advantageous for the secondary recolonization of rainforests (shrub/tree floor) in Albertine Rift Mountains and part of Guineo-Congolian region by the clade of (*Thamnomys* + *G. poensis* group) (sensu Bryja et al., 2017; see below for proposed taxonomic changes).

4.4 Taxonomic implications, delimitation of genera in Arvicanthini

The resolved phylogeny of the tribe provides the opportunity to revise its generic classification. There are neither rules nor generally accepted consensus about what the mammalian genus

should be (Dubois, 1988), contrary to numerous species concepts (e.g. Zachos, 2016), but at least it should consist of a monophyletic group of species characterized by synapomorphic traits. There are at least two cases, where this is not true in current taxonomy of Arvicanthini as reported in the recent Handbook of the Mammals of the World (Wilson et al. 2017). First, Missoup et al. (2018) recently performed phylogenetic analysis of one mitochondrial and two nuclear DNA fragments and found that *Hybomys* (sensu Wilson et al., 2017) is a paraphyletic taxon. Our phylogenomic analysis confirmed this finding and we already follow the generic classification proposed by Missoup et al. (2018), i.e. the split of former *Hybomys* into *Hybomys* and *Typomys*.

The second clearly paraphyletic genus is *Grammomys* (Fig. 1), where *G*. *poensis* is a sister taxon to *Thamnomys*, but not to the other *Grammomys* species. Many authors have asserted that the species of *Thamnomys* and *Gramommys* are in the same monophyletic group and separable only at the subgeneric level, while others consider them as two clearly distinct genera (see references in Musser & Carleton 2005). The genus *Thamnomys* represents a poorly documented group of species that are either rare or difficult to collect, and with very restricted distribution ranges in rainforests of the Albertine Rift and eastern Congo Basin (Wilson et al., 2017). Because of unavailability of samples, they were not included in phylogenetic studies until very recently. Bryja et al. (2017) for the first time used *Thamnomys* sequences in their multi-locus study, but its position remained unresolved. Even if their overall phylogeny of Arvicanthini was based on four mitochondrial and five nuclear DNA markers, for *Thamnomys* only two mitochondrial fragments were available (see similar results in Missoup et al., 2018). Here we unambiguously showed that the *poensis* group (sensu Bryja et al., 2017) is much closer to *Thamnomys* than to *Grammomys* (Fig. 1). This relationship is supported also by morphological traits on the skull and teeth (Hutterer and Dieterlen, 1984) and we therefore propose to classify the *poensis* group (with two species

listed in the most recent compendia, *poensis* and *kuru*; e.g. Monadjem et al., 2015; Wilson et al., 2017) as an internal lineage of *Thamnomys*. The *poensis* group and remaining *Thamnomys* diverged ca. 4 Ma, which is comparable with the first intrageneric splits, e.g. between species of *Aethomys* (Fig. 2). The genera *Thamnomys* (including *poensis* group) and *Grammomys* diverged 5.7 Ma, i.e. well before all other intra-generic diversification events in Arvicanthini (Fig. 2). This ancient divergence and the fact that these two taxa can be distinguished by several morphological characters are strong arguments to consider *Thamnomys* and *Grammomys* as distinct genera and not as subgenera. According to this findings, the two sister genera, *Thamnomys* (with most species occupying Congo basin and Albertine Rift Mts.) and *Grammomys* (with the highest diversity in montane and coastal forests of East Africa), are the descendants of lineages that became separated by the split of Guineo-Congolian and Eastern-African forests during the Miocene/Pliocene boundary (see more details in Bryja et al., 2017). This situation is analogous to another widespread murine genus *Praomys* sensu lato (the tribe Praomyini), where most diversity is currently found in tropical Guineo-Congolian forests. Especially the *P. jacksoni* complex shows very similar phylogeographic structure to *Thamnomys* (in the new view, i.e. including the *poensis* group), with very high diversity in Albertine rift Mts. and east-west structure of populations north of the Congo River (Mizerovská et al., 2019). On the other hand, Eastern African montane forests are inhabited by the so-called *P. delectorum* group (often found together with *Grammomys* species) that diverged very early during the Praomyini radiation in Miocene/Pliocene boundary and should be excluded from the genus *Praomys* (Missoup et al., 2012).

The genomic differences among currently recognized sister genera of Arvicanthini, reflected as their divergence time (Fig. 2), are very variable. For example the evolutionary distances among species within genera *Aethomys*, *Thamnomys* (including the *poensis* group) and *Grammomys*

(even after exclusion of the *poensis* group) are comparable with many intergeneric differences (e.g. *Stochomys*/*Dephomys*, *Lamottemys*/*Desmomys*/*Rhabdomys*) or much higher (*Mylomys*/*Pelomys*, or *Arvicanthis*/*Lemniscomys*). If we assume that most of loci used in this study are selectively neutral, this implies that the extent of genomic differences (and divergence times) is much higher within some genera than between some others. We do not advocate here the split of genetically heterogeneous genera (or the lumping of genetically similar genera), but the outputs of our phylogenomic analysis provide interesting hypotheses worthy of testing by future integrative taxonomic work. They should become a matter for discussions in the mammalogical community and might challenge the present generic classification of Arvicanthini.

Acknowledgements

This study was supported by the project of the Czech Science Foundation, no. 18-17398S, and UMR7205 of the MNHN. We also thank the SYNTHESYS programme (BE-TAF-5113 to OM and FRTAF-5799 to JB). For permission to carry out the research and to collect specimens we are obliged to the Kenyan Forest Service and the Kenyan Wildlife Service, the Ethiopian Wildlife Conservation Authority, the National Research Council and Forestry Department in Malawi, Sokoine University of Agriculture in Morogoro (Tanzania), the National Directorate for Protected Areas (DINAC – Mozambique), and Zambian Wildlife Authority, the Moroccan Haut Commissariat aux Eaux et Forêts et à la Lutte Contre la Desertification, the Benin Direction des forêts et des ressources naturelles from Benin, the Guinean Direction Nationale des Eaux et Forêts and the Ivorian Ministère de l'Agriculture et des Ressources Animales. We would like to thank F. Catzeflis, L. Granjon and J. Kerbis Peterhans for providing four tissue samples. For help in the field and with logistics, we acknowledge S. Gryseels, Y. Meheretu, K. Welegerima, A. Ribas, P. Kaňuch, H. Konvičková, F. Sedláček, R. Makundi, A. Massawe, G. Phamphi, J. Mbau, J. Šklíba, M. Lövy, A. D.

Missoup, and all local collaborators. The authors thank M. Kortyna, S. Holland, K. Birch, and A. Bigelow for assistance with generating the anchored phylogenomics data set. Access to the National Grid Infrastructure MetaCentrum provided under the programme CESNET LM2015042 is greatly appreciated and so is the access to CIPRES Science Gateway (https://www.phylo.org/). These facilities were used to run most of the phylogenetic analyses.

Author contributions

J.B., R.S., A.K., V.N., C.D., and E.V. conceived the study, provided samples and funding, A.R.L. and E.M.L. produced the anchored phylogenomics dataset, A.B. did part of the lab work and assembled mitogenomes, O.M. and J.B. analysed data and drafted the manuscript. All authors contributed to the editing of the manuscript, gave final approval for publication and agreed to be held accountable for the work performed therein.

Data availability

Complete mitochondrial genomes are available in SM3 and in GenBank under accession numbers MN807579-MN807618 (see Table S1 in SM1). Alignments of nuclear loci obtained by anchored phylogenomic approach (as partitioned nexus file) and the Bayesian gene trees used as input for ASTRAL analysis (in newick format) are available in the public repository of the Czech Academy of Sciences (http://hdl.handle.net/11104/0312390) and in Mendeley Data repository (DOI: 10.17632/72cs8jnkg3.1).

Appendix A. Supplementary material

Supplementary Materials SM1 (Table S1): List of used specimens.

Supplementary Materials SM2: The template of MrBayes block specifying the inference of gene trees in the *nexus* file.

Supplementary Materials SM3: The mitogenomic dataset, and MrBayes block specifying the best partition scheme for mtDNA and substitution models in the *nexus* file.

References

Aghová, T., Kimura, Y., Bryja, J., Dobigny, G., Granjon, L., Kergoat, G.J., 2018. Fossils know it best: using a new set of fossil calibrations to improve the temporal phylogenetic framework of murid rodents (Rodentia: Muridae). Mol. Phylogenet. Evol. 128, 98–111.

Amaral, C.R.L., Pereira, F., Silva, D.A., Amorim, A., de Carvalho, E.F., 2018. The mitogenomic phylogeny of the Elasmobranchii (Chondrichthyes). Mitochondrial DNA Part A 29, 867–878.

Bonnefille, R., Potts, R., Chalie, F., Jolly, D., Peyron, O., 2004. High-resolution vegetation and climate change associated with Pliocene *Australopithecus afarensis*. PNAS 101, 12125–12129.

Bryja, J., Šumbera, R., Kerbis Peterhans, J.C., Aghová, T., Bryjová, A., Mikula, O., Nicolas, V., Denys, C., Verheyen, E., 2017. Evolutionary history of the thicket rats (genus *Grammomys*) mirrors the evolution of African forests since late Miocene. J Biogeogr. 44, 182–194.

Bryja, J., Mikula, O., Šmbera, R., Meheretu, Y., Aghová, T., Lavrenchenko, L.A., Mazoch, V., Oguge, N., Mbau, J.S., Welegerima, K., Amundala, N., Colyn, M., Leirs, H., Verheyen, E., 2014. Pan-African phylogeny of *Mus* (subgenus *Nannomys*) reveals one of the most successful mammal radiations in Africa. BMC Evol. Biol. 14, 256.

Castresana, J., 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol. Biol. Evol. 17, 540–552.

Cerling, T.E., Harris, J.M., MacFadden, B.J., Leakey, M.G., Quade, J., Eisenmann, W., Ehleringer, J.R., 1997. Global vegetation change through the Miocene/Pliocene boundary. Nature 389, 153–158.

Degnan, J.H., Rosenberg, N.A., 2009. Gene tree discordance, phylogenetic inference and the multispecies coalescent. Trends Ecol. Evol. 24, 332–340.

deMenocal, P.B., 1995. Plio-Pleistocene African climate. Science 270, 53–59.

Denys, C., Missoup, A.D., Nicolas, V., Fülling, O., Delapré, A., Bilong Bilong, C.F., Taylor, P.J.J., Hutterer R., 2014. African highlands as mammal diversity hotspots: new records of *Lamottemys okuensis* (Rodentia: Muridae) and other endemic rodents from Mount Oku, Cameroon. Zoosystema 36, 647–690.

Denys, C., Lecompte, E., Dubois, A., Taylor, P.J., 2017. Diagnoses and contents of new African and Eurasian Murinae (Rodentia, Muridae) tribes. Dumerilia 7, 82–90.

Denys, C., Winkler, A., 2015. Advances in integrative taxonomy and evolution of African murid rodent: how morphological trees hide the molecular forest. In Cox, P., Hautier, L. (eds.): Evolution of the Rodents: Advances in Phylogeny, Palaeontology and Functional Morphology. Cambridge studies in morphology and molecules. New paradigms in Evolutionary Biology, Cambridge University Press, 186–220.

Drummond, A.J., Bouckaert, R.R., 2015. Bayesian evolutionary analysis with BEAST. Cambridge University Press.

Drummond, A.J., Ho, S.Y.W., Phillips, M.J., Rambaut, A., 2006. Relaxed phylogenetics and dating with confidence. PLoS Biology 4, e88.

Dubois, A. 1988. The genus in zoology: A contribution to the theory of evolutionary systematics. Mém. Mus. Natn. Hist. Nat. (A) 140, 1–124.

Ducroz, J.F., Volobouev, V., Granjon, L., 2001. An assessment of the systematics of arvicanthine rodents using mitochondrial DNA sequences: evolutionary and biogeographical implications. J. Mammal. Evol. 8, 173–206.

Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research 32, 1792–1797.

Feakins, S.J., deMenocal, P.B., 2010. Global and African regional climate during the Cenozoic. In Werdelin, L., Sanders, W.J. (eds): Cenozoic Mammals of Africa. Berkeley, University of California Press, 45–55.

Geraads, D., 2001. Rongeurs du Miocene superieur de Chorora (Ethiopie): Dendromuridae, Muridae et conclusions. Palaeovertebrata 30, 89–109.

Gernhard, T., 2008. The conditioned reconstructed process. J. Theor. Biol. 253, 769–778.

Gissi, C., Reyes, A., Pesole, G., Saccone, C., 2000. Lineage-specific evolutionary rate in mammalian mtDNA. Mol. Biol. Evol. 17, 1022–1031.

Hamilton, C.A., Lemmon, A.R., Lemmon, E.M., Bond, J.E., 2016. Expanding Anchored Hybrid Enrichment to resolve both deep and shallow relationships within the spider Tree of Life. BMC Evol. Biol. 16, 212.

Höhna, S., Landis, M.J., Heath, T.A., Boussau, B., Lartillot, N., Moore, B.R., Huelsenbeck, J.P., Ronquist, F., 2016. RevBayes: Bayesian phylogenetic inference using graphical models and an interactive model-specification language. Syst. Biol. 65, 726–736.

Huelsenbeck, J.P., Larget, B., Alfaro, M.E., 2004. Bayesian phylogenetic model selection using reversible jump Markov chain Monte Carlo. Mol. Biol. Evol. 21, 1123–1133.

Hutterer, R., Dieterlen, F. 1984. Zwei neue Arten der Gattung *Grammomys* aus Äthiopien und Kenia (Mammalia: Muridae). Stuttgarter Beiträge zur Naturkunde Serie A (Biologie) 374, 1–18.

Jacobs, L.L., Flynn, J.J., 2005. Of Mice…again: the Siwalik rodent record, murine distribution, and molecular clocks. In Lieberman, D.E., Smith, R.J., Keller, J. (eds.): Interpreting the Past Essays on Human, Primate, and Mammal Evolution in Honor of David Pilbeam. Brill Academic Publishers, 63–80.

Katoh, K., Toh, H., 2008. Improved accuracy of multiple ncRNA alignment by incorporating structural information into a MAFFT-based framework. BMC Bioinformatics 9, 212

Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. 30, 772–780.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P., Drummond, A., 2012. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28, 1647–1649.

Kimura, Y., Hawkins, M.T.R., McDonough, M.M., Jacobs, L.L., Flynn, L.J., 2015. Corrected placement of *Mus–Rattus* fossil calibration forces precision in the molecular tree of rodents. Sci. Rep. 5, 14444.

Krásová, J., Mikula, O., Mazoch, V., Bryja, J., Říčan, O., Šumbera, R., 2019. Evolution of the Grey-bellied Pygmy Mouse group: highly structured molecular diversity with predictable geographic ranges but morphological crypsis. Mol. Phylogenet. Evol. 130, 143–155.

Lanfear, R., Frandsen, P.B., Wright, A.M., Senfeld, T., Calcott, B., 2016. PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. Mol. Biol. Evol. 34, 772–773.

Larsson, A., 2014. AliView: a fast and lightweight alignment viewer and editor for large data sets. Bioinformatics 30, 3276–3278.

Lavrenchenko, L.A., Verheyen, E., 2005. An assessment of the systematics of the genus *Desmomys* Thomas, 1910 (Rodentia: Muridae) using mitochondrial DNA sequences. In Huber, B.A., Sinclair, B.J., Lampe, K.H. (eds.): African Biodiversity: Molecules, Organisms, Ecosystems. Springer, New York, 363–369.

Lecompte, E., Aplin, K., Denys, C., Catzeflis, F., Chades, M., Chevret, P., 2008. Phylogeny and biogeography of African Murinae based on mitochondrial and nuclear gene sequences, with a new tribal classification of the subfamily. BMC Evol. Biol. 8, 199.

Lemmon, A.R., Emme, S., Lemmon, E.M., 2012. Anchored hybrid enrichment for massively high-throughput phylogenetics. Syst. Biol. 61, 721–744.

Lemmon, E.M., Lemmon A.R., 2013. High-Throughput genomic data in systematics and phylogenetics. Ann. Rev. Ecol. Evol. Syst. 44, 99–121.

Maley, J., 1996. The African rain forest – main characteristics of changes in vegetation and climate from the Upper Cretaceous to the Quaternary. Procs. R. Soc. Edinburgh 104B, 31–73.

Manthi, F.K., 2007. A preliminary review of the rodent fauna from Lemudong'o, south-western Kenya, and its implication to the late Miocene paleoenvironments. Kirtlandia 56, 92–105.

McCormack, J.E., Faircloth, B.C., Crawford, N.G., Gowaty, P.A., Brumfield, R.T., Glenn, T.C., 2012. Ultraconserved elements are novel phylogenomic markers that resolve placental mammal phylogeny when combined with species tree analysis. Genome Res. 22, 746–754.

McCormack, J.E., Tsai, W.L., Faircloth, B.C., 2016. Sequence capture of ultraconserved elements from bird museum specimens. Mol. Ecol. Resour. 16, 1189–1203.

Mirarab, S., Warnow, T., 2015. ASTRAL-II: coalescent-based species tree estimation with many hundreds of taxa and thousands of genes. Bioinformatics 31, i44–i52.

Missoup, A.D., Nicolas, V., Wendelen, W., Keming, E., Bilong Bilong, C.F., Couloux, A., Atanga, E., Hutterer, R., Denys, C., 2012. Systematics and diversification of *Praomys* species (Rodentia:

Muridae) endemic to the Cameroon Volcanic Line (West Central Africa). Zool. Scri. 41, 327–345.

Missoup, A.D., Nicolas, V., Eiseb, S., Chung, E.K., Denys, C., 2016. Phylogenetic position of the endemic Mount Oku rat, *Lamottemys okuensis* (Rodentia: Muridae), based on molecular and morphological data. Zool. J. Lin. Soc. 177, 209–226.

Missoup, A.D., Yemchui, G.D., Denys, C., Nicolas, V., 2018. Molecular phylogenetic analyses indicate paraphyly of the genus *Hybomys* (Rodentia: Muridae): Taxonomic implications. J. Zool. Syst. Evol. Res. 56, 444–452.

Mizerovská, D., Nicolas, V., Demos, T.C., Akaibe, D., Colyn, M., Denys, C., Kaleme, P., Katuala, P., Kennis, J., Kerbis Peterhans, J.C., Laudisoit, A., Missoup, A.D., Šumbera, R., Verheyen, E., Bryja, J., 2019. Genetic variation of the most abundant forest-dwelling rodents in Central Africa (*Praomys jacksoni* complex): Evidence for Pleistocene refugia in both montane and lowland forests. J Biogeogr. 46, 1466–1478.

Monadjem, A., Taylor, P.J., Denys, C., Cotterill, F.P.D., 2015. Rodents of Sub-Saharan Africa. A biogeographic and taxonomic synthesis. Walter de Gruyter GmbH, Berlin/Munich/Boston.

Morley, R.J., Kingdon, J. (eds.), 2013. Mammals of Africa. Volume I: Introductory Chapters and Afrotheria. Bloomsbury.

Musser, G.G., Carleton, M.D., 2005. Superfamily Muroidea. In Wilson, D.E., Reeder, D.A.M. (eds.): Mammal Species of the World. A Taxonomic and Geographic Reference. The Johns Hopkins University Press, Baltimore, 894–1531.

Ogilvie, H.A., Bouckaert, R.R., Drummond, A.J., 2017. StarBEAST2 brings faster species tree inference and accurate estimates of substitution rates. Mol. Biol. Evol. 34, 2101–2114.

Paradis, E., Schliep, K., 2018. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics 35, 526–528.

Patnaik R., 2014. Phylogeny of Siwalik murine rodents: Implications for *Mus-Rattus* divergence time. J. Paleontol. Soc. India 59, 15–28.

Potts R., 2013. Hominin evolution in settings of strong environmental variability. Quat. Sci. Rev. 73, 1–13.

Pozzi, L., Hodgson, J.A., Burrell, A.S., Sterner, K.N., Raaum, R.L., Disotell, T.R., 2014. Primate phylogenetic relationships and divergence dates inferred from complete mitochondrial genomes. Mol. Phylogenet. Evol. 75, 165–183.

Prum, R.O., Berv, J.S., Dornburg, A., Field, D.J., Townsend, J.P., Lemmon, E.M., Lemmon, A.R., 2015. A fully resolved, comprehensive phylogeny of birds (Aves) using targeted next generation DNA sequencing. Nature 526, 569–573.

R Core Team, 2019. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Rambaut, A., Drummond, A.J., 2007. Tracer v1.5, Available from http://beast.bio.ed.ac.uk/Tracer

Rögl, F., 1999. Mediterranean and Paratethys. Facts and hypotheses of an Oligocene to Miocene paleogeography (short overview). Geol. Carpathica 50, 339–349.

Rokyta, D.R., Lemmon, A.R., Margres, M.J., Arnow, K., 2012. The venom-gland transcriptome of the eastern diamondback rattlesnake (*Crotalus adamanteus*). BMC Genomics 13, 312.

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D.L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M.A., Huelsenbeck, J.P., 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst. Biol. 61, 539–542.

Rowe, K.C., Reno, M.L., Richmond, D.M., Adkins, R.M., Steppan, S.J., 2008. Pliocene colonization and adaptive radiations in Australia and New Guinea (Sahul): Multilocus systematics of the old endemic rodents (Muroidea: Murinae). Mol. Phylogenet. Evol. 47, 84–101.

Ruane, S., Raxworthy, C.J., Lemmon, A.R., Lemmon, E.M., Burbrink, F.T., 2015. Comparing large anchored phylogenomic and small molecular datasets for species tree estimation: An empirical example using Malagasy pseudoxyrhophiine snakes. BMC Evol. Biol. 15, 221.

Schenk, J.J., Rowe, K.C., Steppan, S.J., 2013. Ecological opportunity and incumbency in the diversification of repeated continental colonizations by muroid rodents. Syst. Biol. 62, 837–864.

Schliep, K.P., 2011. phangorn: phylogenetic analysis in R. Bioinformatics 27, 592–593.

Ségalen, L., Lee-Thorp, J., Cerling, T., 2007. Timing of C4 grass expansion across sub-Saharan Africa. J. Hum. Evol. 53, 549–559.

Stamatakis, A., Hoover, P., Rougemont, J., 2008. A rapid bootstrap algorithm for the RAxML Web servers. Syst. Biol. 57, 758–771.

Stamatakis, A., 2014. RAxML version 8 a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30, 1312–1313.

Steppan, S., Adkins, R., Anderson, J., 2004. Phylogeny and divergence-date estimates of rapid radiations in muroid rodents based on multiple nuclear genes. Syst. Biol. 53, 533–553.

Steppan, S.J., Adkins, R.M., Spinks, P.Q., Hale, C., 2005. Multigene phylogeny of the Old World mice, Murinae, reveals distinct geographic lineages and the declining utility of mitochondrial genes compared to nuclear genes. Mol. Phylogenet. Evol. 37, 370–388.

Steppan, S.J., Schenk, J.J., 2017. Muroid rodent phylogenetics: 900-species tree reveals increasing diversification rates. PLoS One 12, e0183070.

Šumbera, R., Krásová, J., Lavrenchenko, L.A., Mengistu, S., Bekele, A., Mikula, O., Bryja, J., 2018. Ethiopian highlands as a cradle of the African fossorial root-rats (genus *Tachyoryctes*), the genetic evidence. Mol. Phylogenet. Evol. 126, 105–115.

Talavera, G., Castresana, J., 2007. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. Syst. Biol. 56, 564–577.

Taylor, P.J., Maree, S., Cotterill, F.P.D., Missoup, A.D., Nicolas, V., Denys, C., 2014. Molecular and morphological evidence for a Pleistocene radiation of laminate-toothed rats (*Otomys*: Rodentia) across a volcanic archipelago in equatorial Africa. Biol. J. Linn. Soc. 113, 320–344.

Tucker, D.B., Colli, G.R., Giugliano, L.G., Hedges, S.B., Hendry, C.R., Lemmon, E.M., Lemmon, A.R., Sites, J.W.Jr., Pyron, R.A., 2016. Phylogenomic analysis of tegus and whiptails (Teiidae: Squamata), with a revised taxonomy and a new genus from the West Indies. Mol. Phylogenet. Evol. 103, 75–84.

Wilson, D.E., Lacher, T.E.Jr., Mittermeier, R.A. (eds), 2017. Handbook of the Mammals of the World, vol. 7. Rodents II. Lynx Editions, Barcelona.

Winkler, A.J., 2002. Neogene paleobiogeography and East African paleoenvironments: contributions from the Tugen Hills rodents and lagomorphs. J. Hum. Evol. 42, 237–256.

Winkler, A.J., Denys, C., Avery, M.D., 2010. Rodentia. In Werdelin, L., Sanders, W.J. (eds): Cenozoic Mammals of Africa. Berkeley, University of California Press, 263–304.

WoldeGabriel, G., White, T.D., Suwa, G., Renne, P., de Heinzelin, J., Hart, W.K., Heiken, G., 1994. Ecological and temporal placement of early Pliocene hominids at Aramis, Ethiopia. Nature 371, 330–333.

Wynn, J.G., Alemseged, Z., Bobe, R., Geraads, D., Reed, D., Roman, D.C., 2006. Geological and palaeontological context of a Pliocene juvenile hominin at Dikika, Ethiopia. Nature 443, 332–336.

Zachos, F., 2016. Species Concepts in Biology. Historical Development, Theoretical Foundations and Practical Relevance. Springer International Publishing, Switzerland.

Figure legends

Figure 1 Species tree based on 377 nuclear loci (from ASTRAL; in total 581,030 bp) and Bayesian estimate of mitochondrial phylogeny (from MrBayes; 14,777 bp). Nodes are coloured according to posterior probabilities from ASTRAL and MrBayes (squares) and bootstrap support from the maximum likelihood (RAxML) analyses (circles). The colours distinguish categories of statistical support.

Figure 2 Divergence dating of the species tree inferred using a multi-species coalescent approach in StarBEAST2. The analysis were based on 231 loci from the anchored phylogenomic dataset and the molecular clock was calibrated by two fossil constraints (the root and the MRCA of *Aethomys* and *Arvicanthis*). The numbers in circles are TMRCAs of particular clades in million years ago (Ma).

Highlights

- fully resolved phylogeny of a highly diversified group of African mammals
- comparison of "anchored phylogenomics" and mitogenomics
- mechanisms of adaptive radiation starting in the Messinian stage (ca. 7 Ma)
- delimitation of genera in Arvicanthini and corresponding taxonomic changes

Authors´ Statement

None of the material in this manuscript has been published or is under consideration for publication elsewhere. All data used in the paper are made available. The manuscript has been approved by all the co-authors who agreed to its submission.

Thank you for considering this manuscript, we will look forward to hearing from you.

Yours Sincerely,

Josef Bryja (on behalf of all co-authors)